THE COSMIC DISTANCE LADDER

17.1 Introduction

In 1929, the Carnegie astronomer Edwin Hubble, working with data obtained at the Mount Wilson Observatory in California, published the first plot showing that galaxies are receding from us with a radial velocity proportional to their distance. The logical conclusion is that our Universe is in a state of expansion, and Hubble's discovery stands as one of the most profound of the twentieth century. This result had been anticipated two years earlier by Lemaître, who found a mathematical solution for an expanding universe and noted that it provided a natural explanation for the observed receding velocities of galaxies. His results were published (in French) in the Annals of the Scientific Society of Brussels and were not widely known until 1931 when they were highlighted by Eddington.

As can be appreciated from inspection of Figure 17.1, the slope of the original 'Hubble diagram' implies a value of the Hubble constant, $H_0 = v/d$ which is one order of magnitude greater than the value generally accepted today, $H_0 \simeq 70\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ with an uncertainty of about 10%. While it is relatively straightforward to measure the recession velocities of galaxies

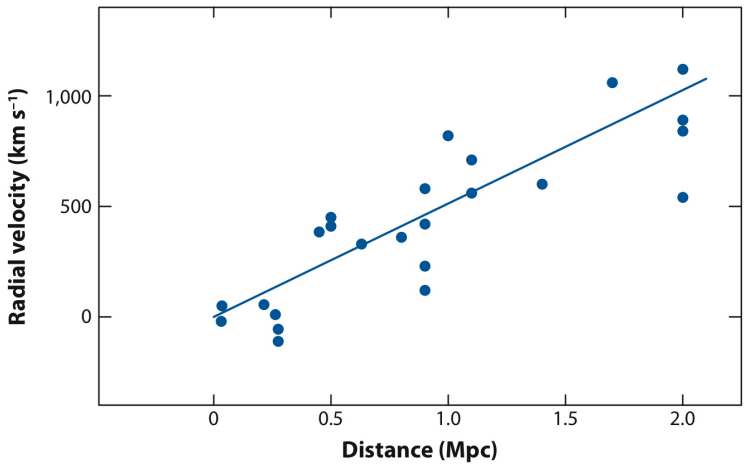

Figure 17.1: The original Hubble diagram, reproduced from Freedman & Madore 2010, ARAA, 48, 673.

from the redshifts of emission and/or absorption lines in their spectra, the determination of distances to astronomical objects is fraught with difficulties and plagued by systematic uncertainties.

Indeed, the history of the determination of the Hubble constant does not show astronomers in a good light: for decades there were two 'camps', one claiming $H_0 = 50\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ and the other double that value; both groups estimated their error to be about 10% and were generally unwilling to concede that systematic uncertainties could significantly inflate their error estimates. The issue was finally resolved in the mid-1990s, largely thanks to observations performed with the *Hubble Space Telescope*. Given the central importance of an accurate measure of the Hubble constant for the whole of cosmology, you will not be surprised to learn that this was one of the original scientific motivations for building the *Hubble Space Telescope* and that a great deal of observing time was devoted to this *'Key Project'* in the first few years of operation of *HST*.

Measuring the Hubble constant requires constructing what is referred to a 'cosmic distance ladder'. Distances to objects at cosmological distances are established step by step, identifying classes of astronomical sources who act as 'standard candles' over ever increasing distances. The calibration of one

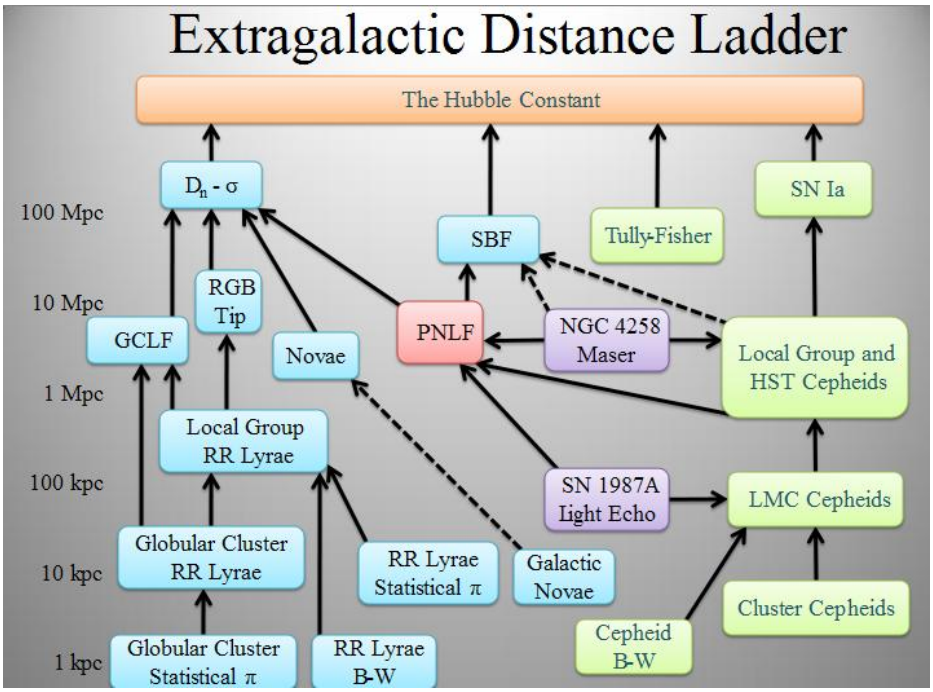

Figure 17.2: The cosmic distance ladder.

class of sources at nearby distances, for example a particular type of star, is used to calibrate the intrinsic luminosity of another type of source, which being intrinsically more luminous (but also rarer), can then be followed to larger distances, and so on. Clearly systematic errors can build very quickly!

In general terms, to measure H_0 distance measurements must be obtained far enough away to probe the smooth Hubble expansion (that is, at sufficiently large distances that the random velocities induced by gravitational interactions with neighboring galaxies are small relative to the Hubble velocity), and nearby enough to calibrate the absolute, not simply the relative, distance scale. The objects under study also need to be sufficiently abundant that their statistical uncertainties do not dominate the error budget. Ideally the method should have a solid physical underpinning and high internal accuracy, amenable to empirical tests for systematic errors.

Over the years, many classes of astronomical sources have been explored with a view to establishing their suitability as standard candles (see Figure 17.2). Here we shall briefly review some of the main ones.

17.2 Cepheids

Chepeids are the first rung in the extragalactic distance ladder and hold a special place in the subject because they were used by Hubble in his first radial velocity vs. distance plot that led to the discovery of the expanding Universe. Cepheids are a class of variable stars located in the upper H-R diagram (see Figure 17.3); they are evolved, core helium-burning stars whose progenitors are thought to have been B- or late O-type main sequence stars. Their visual magnitudes vary in a regular fashion with amplitudes of between a few tenths of magnitude and ~ 2 magnitudes, with periods ranging from a few days to a few weeks (see right-hand panel of Figure 17.3).

The key role that Cepheids have played in the determination of the extragalactic distance scale stems from the existence of a tight period-luminosity relation: the longer the period of their variability, the brighter their absolute magnitude. Cepheids are supergiant stars with $R \sim 50\,\mathrm{R}_\odot$ and $L \gtrsim 10^3\,\mathrm{L}_\odot$, bright enough to be seen over intergalactic distances. Thus, by measuring the period of an extragalactic Cepheid star, it is possible to

deduce its distance modulus by comparing its observed magnitude with the absolute magnitude, *provided the period-luminosity relationship has been calibrated with the known distances of nearby Cepheids in the Milky Way*. A later refinement of the calibration includes a colour term, giving the period-luminosity-colour (PLC) relation:

$$M_{\rm V} = \alpha \log P + \beta (B-V)_0 + \gamma$$

where α is a negative number and $(B-V)_0$ is the intrinsic $B-V$ colour obtained after correcting for interstellar extinction.

Cepheids stars undergo periodic radial pulsations that are at the root of the PLC relation via Stephan's law: $L = 4\pi R^2 \sigma T^4$. The change in luminosity is driven primarily by changes in surface temperature, while the period is related to the stellar radius via the mean density (for mechanical systems it is well known that $P\rho^{1/2} = Q$, where Q is a structural constant).

Cepheid pulsation occurs because of the changing atmospheric opacity with temperature in the helium ionization zone (the zone where He transitions from He^+ to He^{2+}). This zone acts like a heat engine and valve mechanism. During the portion of the cycle when the ionization layer is opaque to radiation, that layer traps energy, resulting in an increase in its internal pressure. This added pressure acts to elevate the layers of gas above it,

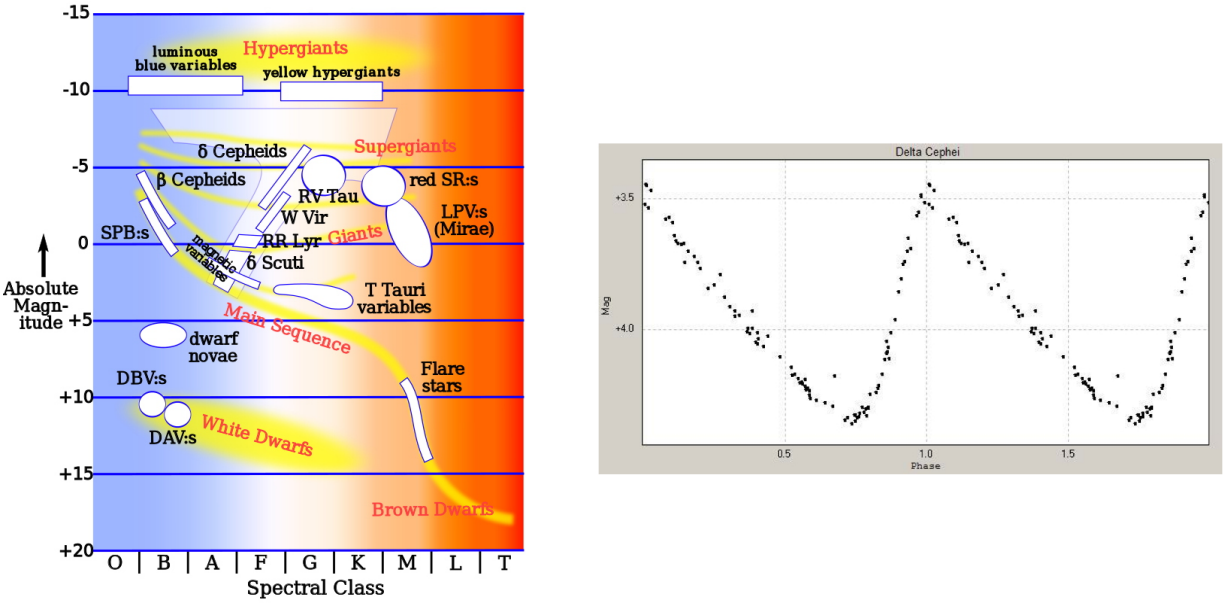

Figure 17.3: *Left:* Locations of different types of variable stars in the H-R diagram. *Left:* Light curve of δ Cephei, the prototype of the Cepheid variable stars.

resulting in the observed radial expansion. As the star expands, it does work against gravity and the gas cools. As it does so, its temperature falls back to a point where the doubly ionized helium layer recombines and becomes transparent again, thereby allowing more radiation to pass. Without that added source of heating the local pressure drops, the expansion stops, the star recollapses, and the cycle repeats. The alternate trapping and releasing of energy in the helium ionization layer ultimately gives rise to the periodic change in radius, temperature, and luminosity seen at the surface. Not all stars are unstable to this mechanism. The cool (red) edge of the Cepheid instability strip is thought to be controlled by the onset of convection, which then prevents the helium ionization zone from driving the pulsation. For hotter temperatures, the helium ionization zone is located too far out in the atmosphere for significant pulsations to occur.

There are complications, of course. Stars spend only a short fraction of their lives in the so-called 'instability strip' of the H-R diagram. Therefore, Cepheids are relatively rare and we have to go a long way from the Sun before we encounter one. Consequently, there are just *ten* Cepheids with accurately measured trigonometric parallaxes.[1] The resulting error in the Cepheid zero point is ± 0.06 mag ($\pm 3\%$). This error will be reduced considerably when data from the on-going *Global Astrometric Interferometer for Astrophysics* (GAIA) mission are gathered and analysed.

Further complications are: interstellar reddening, which can be mitigated by observations at infrared wavelengths; crowding of stellar images (this is where *HST* played a key role with its superior spatial resolution); and the dependence of the PLC relation on the metallicity of the stars. Concerning this last point, it should be clear from your *Stellar Structure and Evolution* course that metallicity has an impact on the opacity of stellar layers. Since it is changes in opacity that drive stellar pulsations, it is reasonable to expect that there will be a metallicity dependence of the PLC relation. However, predicting the magnitude (and even simply the sign of the effect) at either optical or infrared wavelengths, has proved challenging. Different theoretical and empirical studies have led to a range of conclusions, and the effect of metallicity on the observed properties of Cepheids is still an active and on-going area of research.

[1] Other methods to calibrate the Cepheid zero point have been proposed, but none is as reliable as accurately measured parallaxes.

17.3 The Distance to the Large Magellanic Cloud

The next rung in the cosmic distance ladder, and the first one outside the Milky Way, is our nearest companion galaxy, the Large Magellanic Cloud (LMC). Given its importance as a stepping stone to extragalactic distances and its proximity to our Galaxy, there have been numerous attempts at measuring the distance to the LMC.

One of the most direct uses the light echoes of SN 1987A, the closest supernova since the invention of the telescope. Approximately 240 days after the supernova explosion, a ring of circumstellar material, ejected during an earlier phase in the evolution of the supernova progenitor, became visible as the flash of UV radiation accompanying the SN explosion reached it and ionised it. (This is the inner ring in Figure 17.4; the origin of the two outer rings is still unclear.) For the inner ring, we know its physical radius, $R = c\,\Delta t$ with $\Delta t = 240\,\mathrm{days}$, and its angular radius on the sky, $\theta = 0.85\,\mathrm{arcsec}$,[2] and we can therefore solve for the distance to the SN:

$$D_{\mathrm{LMC}} = \frac{c\Delta t}{\theta} = 1.51 \times 10^{23}\,\mathrm{cm} \;=\; 48.9\,\mathrm{kpc}$$

which corresponds to a distance modulus $(m - M)_{\mathrm{LMC}} = 18.45$.

[2]The ring is actually inclined to our line of sight which is why we see it as an ellipse, but this geometric effect can be accounted for.

Figure 17.4: Images of SN1987A and its immediate surroundings recorded with the *Hubble Space Telescope* on 6 February 1996 (*Left*) and 6 December 2006 (*Right*). The inner ring has been brightening in clumps as the shock front from the supernova began colliding with the slower moving material ejected by the SN progenitor during its red supergiant phase.

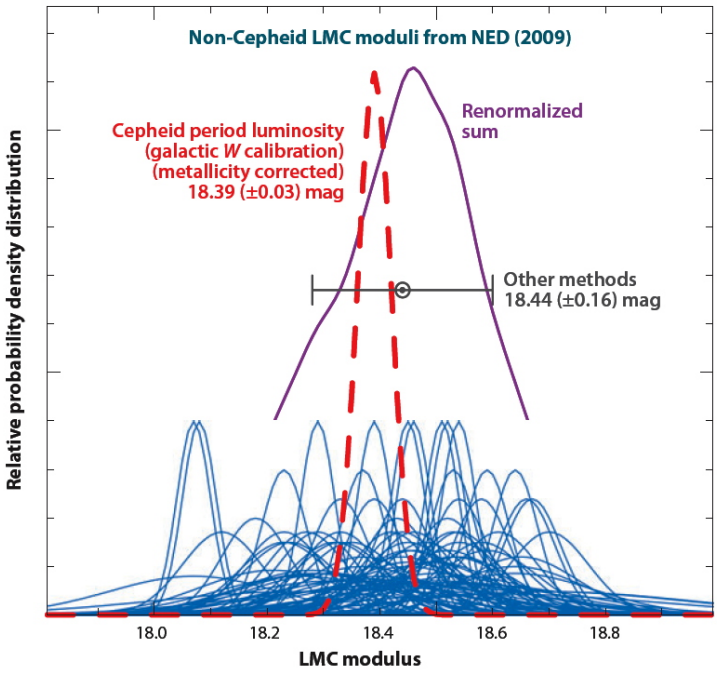

Figure 17.5: Probability distributions of 180 distance modulus estimates to the LMC. Individual estimates are shown by unit-area Gaussians with a dispersion set to their quoted statistical errors (thick blue lines). The thin solid purple line represents the renormalized sum of those Gaussians. The thick broken red line represents the value $(m - M) = 18.39$ mag with a systematic error of ± 0.03 mag based on the Galactic parallax calibration of Cepheids (see Section 17.2) and corrected for the LMC metallicity by -0.05 mag. The circled dot with error bar shows the median value of the non-Cepheid distance determinations. Note that a distance modulus $m - M = 18.40$ corresponds to a distance of 47.9 kpc. (Figure reproduced from Freedman & Madore 2010, ARAA, 48, 673).

Several thousand Cepheids have been identified and cataloged in the LMC, all at essentially the same distance. As can be seen from Figure 17.5, the calibration of the period-luminosity relation based on the trigonometric parallaxes to the ten Galactic Cepheids discussed in Section 17.2, with a small adjustment to allow for the subsolar metallicity of the LMC ($Z_{\rm LMC} \simeq 0.5 Z_\odot$), results in distance modulus to the LMC that is statistically indistinguishable from the value deduced from the combined analysis of the highly heterogeneous set of determinations based on other methods.

17.4 Tip of the Red Giant Branch Method

A completely independent method for determining distances to nearby galaxies, and one that has comparable precision to the Cepheids, is to

measure the magnitude of stars at the tip of the red giant branch in the H-R diagram (referred to as the TRGB method for short). The TRGB method uses the theoretically well-understood and observationally well-defined discontinuity in the luminosity function of stars evolving up the red giant branch in old, metal-poor stellar populations. This feature has been calibrated using Galactic globular clusters, and because of its simplicity and straightforward application, it has been widely used to determine distances to nearby galaxies.

You will recall from the *Stellar Structure and Evolution* course that stars on the RGB are powered by H shell burning above an inert core of He ash. As the core mass increases, the core actually shrinks (a well known property of a degenerate gas), and its temperature increases until it reaches a value high enough to ignite He fusion in a runaway process: the so-called helium flash. Stars at the tip of the RGB are at this stage. Following

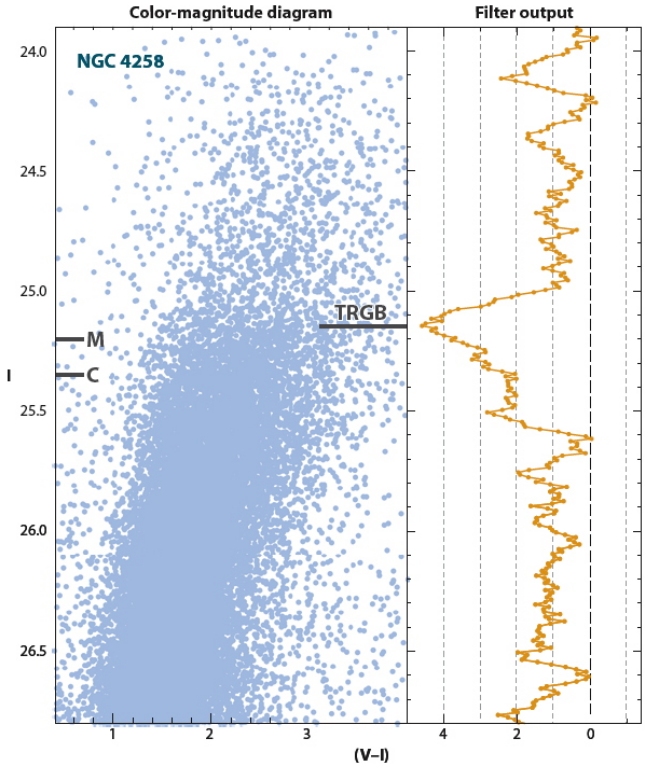

Figure 17.6: An example of the detection and measurement of the TRGB. The left panel is the colour-magnitude diagram of red giant branch stars in the halo of the galaxy NGC 4258. The right panel shows the output of an edge-detection filter (yellow line), whose peak response indicates the TRGB magnitude and whose width is used as a measure of the random error on the detection. (Figure reproduced from Freedman & Madore 2010, ARAA, 48, 673).

the He flash, the star quickly settles on the core He-burning horizontal branch. The transition from the red giant to the horizontal branch occurs rapidly, so that observationally the TRGB can be treated as a physical discontinuity.

The tip of the red giant branch is best identified from colour-magnitude diagrams normally involving the near-infrared I-band, using a quantitative digital filter technique, as illustrated in Figure 17.6.

The advantage of the TRGB method is its observational efficiency because, unlike Cepheid variables, there is no need to follow stars through a variable light cycle: a single-epoch observation through two filters is sufficient. Consequently, there are $\sim 5\times$ more galaxies with distances determined by the TRGB method than via the period-luminosity relation of Cepheids. On the other hand, red giant branch stars are not as bright as Cepheids, and therefore cannot be seen as far. The two methods have been used together for galaxies with distances $d \lesssim 20\,\mathrm{Mpc}$, providing important consistency checks.

17.5 The Maser Galaxy NGC 4258

The acronym Maser stands for Microwave Amplification by Stimulated Emission of Radiation, but a more accurate definition would substitute "Molecular" for "Microwave". The first maser was built in a physics laboratory in the US in the mid 1950s, and it was the precursor to the laser. Masers have also been observed in astrophysical environments, in particular in circumstellar regions, in the interstellar medium near regions of star formation, and in the accretion disks surrounding the massive black holes at the centres of galaxies with Active Galactic Nuclei, where the density of molecules and of infrared photons that can pump their energy levels are both high.

The archetypal AGN megamaser is in the nucleus of the Seyfert 2 galaxy NGC 4258 (a.k.a. M 106) in the Virgo cluster. The maser is detected via the 22 GHz emission line of ortho-H_2O which arises from trace amounts of water vapor ($< 10^{-5}$ in number density) in small density enhancements in in a very thin, slightly warped accretion disk (see Figure 17.7).

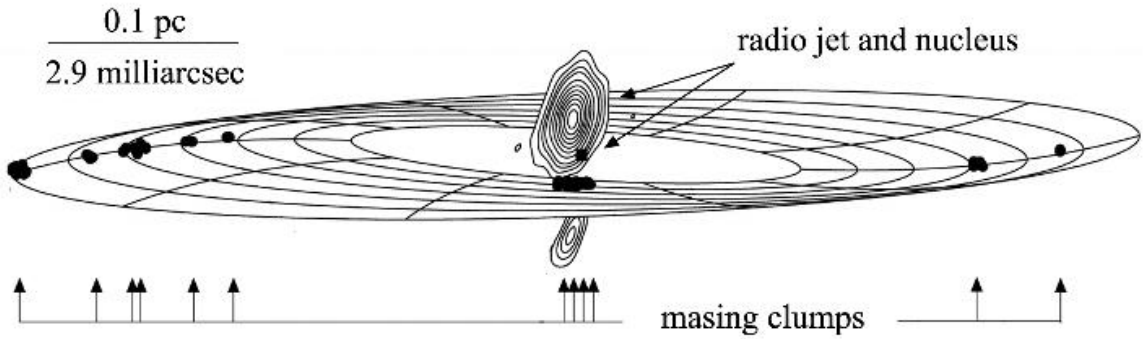

Figure 17.7: Schematic representation of masers in the molecular gas disk in the centre of the Seyfert galaxy NGC 4258. The positions of the masing clumps are shown as they appear on the sky, and superimposed on a model of a nearly edge-on warped disk. Based on their measured Doppler shifts, the velocities of the clumps on the right (approaching) and on the left (receding) of this sketch follow a Keplerian $r^{-1/2}$ curve, indicating a mass of $3.8 \times 10^7\,\mathrm{M}_\odot$ for the central black hole. The proper motions of the clumps near the line of sight to the centre can be tracked over time as they move from right to left on their circular orbits. Comparison of their angular velocities to their physical velocities gives a direct measurement of the distance to the galaxy. (Figure reproduced from D. Maoz, *Astrophysics in a Nutshell*, Princeton University Press).

The masing clumps act as dynamical test particles. A rotation curve is measured along the major axis of the accretion disk by monitoring the change of maser radial velocities over time from single-dish radio observations. Proper motions are measured on the near side of the disk minor axis from observed changes in angular position in VLBI (very long baseline interferometry) images (a very high angular resolution, $\sim 300\,\mu$arcsec) is required for this measurement). A comparison of the angular velocities in the latter measurement with the absolute velocities in km s^{-1} in the former then yields the distance to NGC 4258, $d = 7.2 \pm 0.2\,\mathrm{Mpc}$ (or $(m - M) = 29.29 \pm 0.06\,\mathrm{mag}$). This value is in excellent agreement with the distance to the galaxy deduced from the Cepheids and TRGB methods: $(m - M) = 29.28 \pm 0.04 \pm 0.12\,\mathrm{mag}$.

Attempts to measure distances to other megamasers have proved difficult so far due to lack of sensitivity and the geometric constraint that the maser disk be viewed nearly edge-on. However, the future looks promising for this technique with the planned Square Kilometer Array, due to come into operation in the next decade.

17.6 Tully-Fisher Relation

In 1977, Brent Tully and Richard Fisher found a strong correlation between the total luminosity of a spiral galaxy (corrected to face-on inclination to account for extinction) and the galaxy's maximum (corrected to edge-on inclination) rotation velocity. In a general sense, the TF relation can be understood in terms of the virial relation applied to rotationally supported disk galaxies, under the assumption of a constant mass-to-light ratio. However, a detailed self-consistent physical picture that includes the role of dark matter in producing almost universal spiral galaxy rotation curves still remains a challenge.

The TF relation is at present one of the most widely applied methods for distance measurements, providing distances to thousands of galaxies both in the general field and in groups and clusters. The scatter in this relation is wavelength-dependent, and appears to be greatly reduced at mid-IR wavelengths (see Figure 17.8). If the 3.6 μm correlation stands the test of time as additional calibrators enter the regression, a single galaxy could potentially yield a distance accurate to $\pm 5\%$. All TF galaxies, when observed in the mid-IR, would then individually become precision probes of large-scale structure, large-scale flows, and the Hubble expansion.

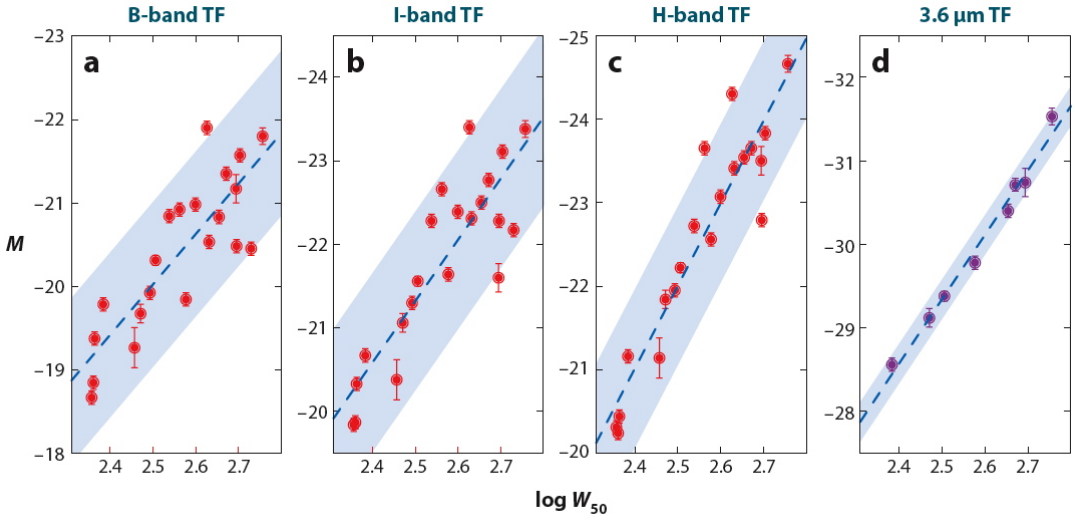

Figure 17.8: Multiwavelength Tully-Fisher (TF) relations for all of the galaxies calibrated with independently measured Cepheid moduli from the *HST* Key Project. Note the significantly reduced dispersion in the mid-IR dataset (panel d); however, a larger sample of calibrators is needed to confirm the scatter and slope of the relation at the wavelength of 3.6 μm. (Figure reproduced from Freedman & Madore 2010, ARAA, 48, 673).

17.7 Type Ia Supernovae

One of the most accurate means of measuring cosmological distances out into the Hubble flow utilizes the peak brightness of SNe Ia. The use of Type Ia SNe as standard candles was already discussed in Lecture 6 and we will not repeat the relevant arguments here.

For Hubble constant determinations, the challenge in using SNe Ia remains that few galaxies in which SN Ia events have been observed are also close enough for Cepheid distances to be measured. Hence, the calibration of the SN Ia distance scale is still subject to small-number statistical uncertainties. At present, the numbers of galaxies for which there are high-quality Cepheid and SN Ia measurements is limited to six objects (see Figure 17.9). Each of the six galaxies has between 13 and 26 Cepheids observed with *HST* in the near-IR H-band.

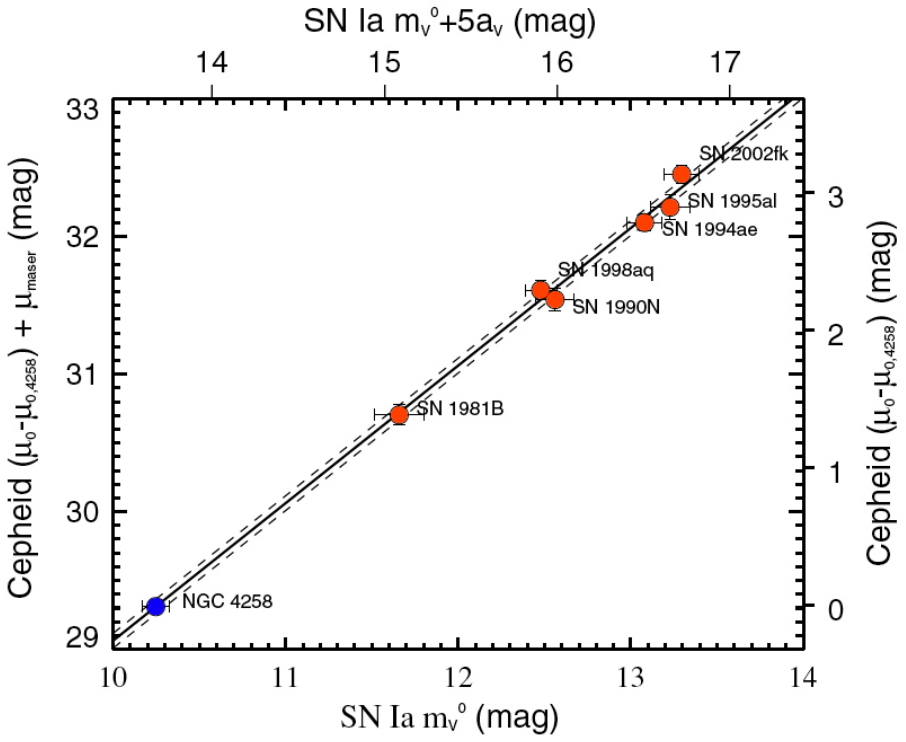

Figure 17.9: A comparison of Cepheid and Type Ia supernovae distances. The authors of this study (Riess et al. 2009) used the maser distance to NGC 4258 to calibrate the Cepheids PL relation, and then adopted NGC 4258 as the calibrating galaxy for the supernova distance scale.

17.8 Other Methods

Among other methods that have been used to determine H_0 we briefly mention the following.

- **Surface Brightness Fluctuations.** In distant galaxies, the unresolved stellar light within a given small solid angle is produced by N stars. The fluctuations in surface brightness between adjacent small areas is Poissonian, i.e. $\sigma \propto 1/\sqrt{N}$. Since the number of stars included per unit solid angle increases with distance to a galaxy, $N \propto d_{\rm gal}^2$, we have $\sigma \propto 1/d_{\rm gal}$. The proportionality constant can be calibrated in nearby galaxies with well-determined distances.

- **Gravitational Lens Time Delay.** The pathlengths to two images of the same source produced by a foreground gravitational mass are different. If the source is variable, such as a quasar or a supernova, the delay in the arrival time of light from one image compared to the other is proportional to H_0^{-1}, and less dependent on other cosmological parameters, such as $\Omega_{\rm m,0}$ and $\Omega_{\Lambda,0}$.

  Initially, the practical implementation of this method suffered from a number of difficulties, the most important being incomplete knowledge of the mass distribution of the lens. However, more recent careful studies of, for example, the quadruple lens system B1608+656, have resulted in improved precision, giving $H_0 = 71 \pm 3\,{\rm km\ s^{-1}\ Mpc^{-1}}$.

- **The Sunyaev-Zel'dovich Effect in X-ray Galaxy Clusters.** We already discussed the S-Z effect in Lecture 10, where it was shown that the S-Z decrement (produced by the redistribution of CMB photons from the Raleigh-Jeans to the Wien side of the blackbody spectrum through inverse Compton scattering off hot electrons in the intracluster medium) is distance independent (see Figure 10.13). Since the measured X-ray flux from a cluster is distance dependent, the combination of CMB and X-ray observations can be used to deduce H_0.

  Again, the accuracy of this method has improved enormously in recent years, with high signal-to-noise, high angular resolution, S-Z images obtained with ground-based interferometric arrays and high-resolution X-ray spectra.

- **Luminosity of Giant H II Regions.** It has been shown that the luminosity of a star-forming galaxy in the Hβ emission line, $L(\mathrm{H}\beta)$, is well correlated with the velocity dispersion, σ, measured from the same spectral feature. While this technique is not yet competitive with the others mentioned in this section, it could (with better observations) be used to probe cosmic expansion to the highest redshifts, since the Hβ emission line is one of the strongest spectral features of star-forming galaxies.
- **Cosmic Microwave Background and Baryonic Acoustic Oscillations.** As we saw in Lectures 10 and 14 respectively, statistical measures of the CMB temperature and polarization fluctuations, and of the large-scale distribution of galaxies encode a number of cosmological parameters. H_0 is involved in combination with $\Omega_{\mathrm{m},0}$ and $\Omega_{\mathrm{baryons},0}$, so that an accurate independent determination of H_0 can help break degeneracies.

17.9 Concluding Remarks

The long-standing debate as to the value of the Hubble constant has largely, but not completely, been resolved in the 21st century. In their comprehensive 2010 review, Freedman & Madore concluded that:

$$H_0 = 73 \pm 2\ (\mathrm{random}) \pm 4\ (\mathrm{systematic})\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$$

(see Figure 17.10). For comparison, the final (2015) analysis of the Planck observations of the CMB gives $H_0 = 67.5 \pm 1\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$. So, the issue is not totally settled, although the two determinations agree to within $\pm 1\sigma$ (if the systematic errors have been estimated correctly). Recently, the geometric maser distance to NGC 4258 has been revised to $7.60 \pm 0.17 \pm 0.15\,\mathrm{Mpc}$ which lowers H_0 by $\sim 3\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$, but also produces some tension with the zero-point of the Galactic Cepheids PL relation. Most works now adopt $h \equiv H_0/100\,\mathrm{km\ s^{-1}\ Mpc^{-1}} = 0.7$ with an accuracy of 5–10%.

Looking ahead, there are still strong motivations for improving further the precision of the determination of the Hubble constant: not only does H_0 set the scale for all cosmological distances and times, but its accurate

determination is also needed to take full advantage of the increasingly precise measurements of other cosmological quantities. The on-going GAIA mission, and the imminent launch of the *James Webb Space Telescope* in particular, are expected to reduce the systematic uncertainty of the zero point of the Cepheid PL relation which underpins many subsequent astronomical distance measures. All of the other methods briefly reviewed in Section 17.8 have undergone major improvements in recent years and will undoubtedly continue to improve in the years ahead.

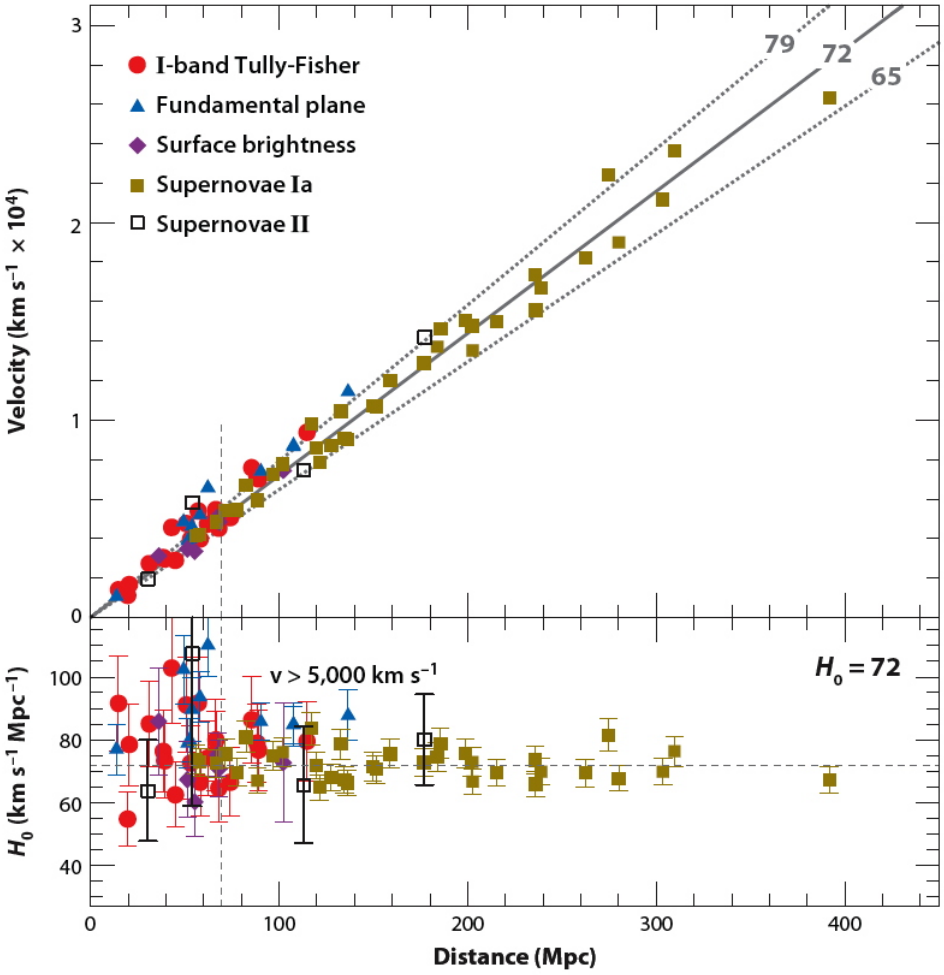

Figure 17.10: Final results from the *Hubble Space Telescope* Key Project to measure the Hubble constant. The lower panel shows the galaxy-by-galaxy values of H_0 as a function of distance. The vertical dashed line is drawn at 5 000 km s^{-1}, beyond which the effects of peculiar motions are thought to be small. (Figure reproduced from Freedman & Madore 2010, ARAA, 48, 673).